МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное

учреждение высшего образования

«Ярославский государственный университет им. П.Г. Демидова»

Математический факультет

Кафедра компьютерной безопасности

Заведующий кафедрой:

д. ф. - м. н., профессор

Мурин Д. М.

Направление подготовки

10.05.01 Компьютерная безопасность

Курсовая работа

Обнаружение аномальной активности в сети на основе анализа статистических отклонений в работе системы.

Научный руководитель

д. ф. - м. н., доцент

\_\_\_\_\_\_\_\_ Д. А. Савинов

«\_\_» \_\_\_\_\_\_\_\_\_ 2025 г.

Студент группы КБ-51СО

\_\_\_\_\_\_\_\_ А.Ю.Сластухин

«\_\_» \_\_\_\_\_\_\_\_\_ 2025 г.

Ярославль 2024 г.

Оглавление

Оглавление 2

Введение 3

IPS/IDS системы 4

Классификация (По области применения) 4

Классификация (По принципу действия) 4

Сигнатурные 4

Аномальные 5

Сферы применения поиска аномалий. 5

Подходы к выявлению аномалий. 6

Прогнозирование временных рядов при решении задачи поиска аномалий. 8

Рекомендательная система как система поиска аномалий. 9

SNORT 10

Своя IDS система 11

Сравнение 12

Литература 14

Введение

IPS/IDS системы

IDS (Intrusion Detection System) - система обнаружения вторжений.

IPS (Intrusion Prevention System) - система предотвращения вторжений.

IDS/IDS является более сложным, комплексный решением, которое может сочетать в себе, анализ файлов, как это делает антивирус, контроль трафика и работу с правилами сетевого взаимодействия, как это делает вайервол.

Классификация (По области применения)

* Network Intrusion Detection System (NIDS) - контролирует работу в сети, взаимодействие нескольких хостовых узлов.
* Host-based Intrusion Detection System (HIDS) – работает на конкретном хосте, не имеет доступа к другим узлам.

NIDS по своей идее очень похожа на вайервол, просматривает весь входящий и исходящий трафик, но является более мощным инструментом, поскольку может находить глубокие признаки аномальной активности.

Основным минусом является тяжеловесность – для глубокого анализа на наличие угроз требуются значительные вычислительные мощности, сопоставимые с мощностями всей системы, поскольку часто требуется обработать всё, что происходит в сети.

Часто NIDS нацелены именно на анализ сетевого взаимодействия в то время, как HIDS анализируют всё, что происходит на конкретном узле, включая трафик, файловую системы, журналы событий – таким образом, NIDS ставят на всю сеть, а на критически важные точки ставят HIDS.

Дополнительно выделяют:

* PIDS (Perimeter Intrusion Detection Systems) - подобно вайерволу контролируют только периметр сети, а не всю сеть.
* VMIDS (Virtual Machine-based Intrusion Detection Systems) – принципиальным отличием является то, что устанавливается на устройство, а на виртуальную машину, что позволяет быстрее её переносить.

Классификация (По принципу действия)

Сигнатурные

Анализируют текущее состояние системы на предмет нахождения уже известных сигнатур угроз. В этом отношении такие системы пересекаются с антивирусами, имеют минус – не реагируют на новые уязвимости.

*Сигнатурой* называется шаблон уже известной угрозы.

Помимо сигнатур выделяют *состояния*: в начальный момент времени считаем, что система в безопасном состоянии, но каждое действие в системе: несанкционированное действие пользователя, сетевое обращение, установка и активация ПО – особенно драйверов, может привести в скомпрометированное состояние. Очевидно, что NIDS часто отслеживают сигнатуры, а HIDS состояния.

Аномальные

*Аномалией* является любое действие, которое хоть по каким-то признакам отличается от нормы, что даёт возможность предполагать, что в система могла перейти в скомпрометированное состояние.

Аномалии делятся на:

* Статистические – сопоставляют профиль штатной работы с текущим, используют статистические алгоритмы. Например, в текущем профиле повышена нагрузка сеть или диск, число операций, интервал между ними – снижены или наоборот завышены. Всё, что указывает на нарушение штатной работы.
* Поведенческие – части системы пытаются выполнять функции, которые им не предусмотрены или нарушается порядок выполнения, например, есть приложение пытается отправить результат до того, как задача была выполнена, или же штатный протокол используется не по назначению.
* Аномалии трафика – статистически трафик может быть штатным, но в нём могут присутствовать нетипичные особенности.

Аномалии могут иметь черты разных типов – это может быть аномалия трафика, в которой помимо угрожающей активности может присутствовать нарушение профиля штатной активности.

В рамках этой работы особый интерес для нас представляют именно системы, работающие с аномалиями. Они более сложные, разнообразные, наукоёмкие с точки зрения используемых подходов, тяжеловесные с точки зрения требуемых ресурсов.

Аномальные системы способны помочь выявить новую угрозу, в то же время они не дают никакой гарантии, что в принципе сработают, в то время как сигнатурные гарантируют, что конкретные сценарии практически недостижимы.

Часто под аномальными представляют нейросети, их действительно можно применять, но только ими дело не ограничивается.

Сферы применения поиска аномалий.

Прежде, чем изучать способы выявления аномалий, рассмотрим несколько сфер возможного применения выявления аномалий:

1.*Анализ штатной активности уже работающих систем* - в 2010 году сетевой червь *Stuxnet* атаковал иранское предприятие, изменив настройки оборудования, привёл к преждевременному выходу из строя центрифуг для обогащения уранового топлива. Есть веские причины полагать, что вирус был разработан при сотрудничестве США и Израиля с целью саботировать иранскую ядерную программу – если бы система контроля учитывала штатную активность оборудования, экономических и стратегических потерь можно было бы избежать.

2.*Выявление аномальных потребительских паттернов или мошеннических действий* – пользователь Боб периодически совершает переводы Алисе. В случае если Боб попытается перевести Алисе сильно больше, чем он делает обычно или попытается сделать перевод в другую страну, хотя обычно он совершает переводы в рамках своего города – банковская система должна отреагировать и запросить дополнительное подтверждение или заблокировать операцию.

3.*Рекоммендательные системы* – бывает очень важно отвечать на вопрос, понравится ли пользователю той или иной продукт, опираясь на профиль его активности – какие категории товаров он смотрел в последнее время, что купил –важно предоставить пользователю связанные товары: купил человек подгузники, вероятно у него есть ребёнок, можно в ленту добавить детское питание, одежду, игрушки.

Возникает вопрос – при чём тут аномалии? А при том, что рекомендательная система может быть преобразована в детектор аномалий.

Например:

* штатная активность пользователя – это основа для рекомендательной системы, аналог просмотрам и покупкам.
* новое действие пользователя – это то, чему система даёт оценку.
* результат работы – если действие может считаться рекомендованным, то штатный профиль активности не нарушен, иначе это аномальная ситуация.

Возникает вопрос, что первично, а что вторично – сложный вопрос, скорее всего, это одна и та же сущность, рассмотренная под разные сферы применения.

Отмечу, что практические вопросы, возникающие при построении рекомендательных систем, такие же трудные, как и проблемы построения систем выявления аномалий – почва для глубоких исследований эффективных подходов.

Подробнее рассмотрим позже.

Приходим к выводу, что поиск аномалий – важная задача.

Подходы к выявлению аномалий.

Важно понимать, что выявление аномалий нетривиальная задача – существует множество алгоритмов, но в большинстве они используют методы, которые описаны ниже.

1.Метод *опорных векторов* с одним классом One-Class SVM

(!) Расписать подробнее

Подходит, когда в обучающем наборе данные подчиняются нормальному распределению, а в тестовом содержат аномалии.

Считается, что этот подход наиболее часто применимый при поиске аномалий.

2. Метод *изолирующего леса* – isolate forest

(!) Расписать подробнее

Опирается то, что при «случайном» построении деревьев выбросы будут попадать в листья на ранних этапах (на небольшой глубине дерева), т.е. выбросы проще «изолировать». Выделение аномальных значений происходит на первых итерациях работы алгоритма.

3. *Elliptic envelope*  и другие статистические методы.

4. *Метрические* методы

5. *Кластерные* методы

6. Метод *главных компонент*

7. Методы на базе *прогнозирования временных рядов* – если рассматриваемое значение выбивается из временного ряда, то значение считается аномальным.

8.Обучение с учителем – регрессия и классификация

9.Методы на базе алгоритмов решения задач рекомендаций.

Решение задачи поиска аномалий часто требует индивидуального подхода к конкретной проблеме – это искусство, требующее своевременного применения различных знаний.

Более того, доказано, что не существует панацеи для борьбы с вирусами, сетевыми угрозами – для любой существующей системы можно предложить подход, позволяющий обойти алгоритмы детектирования. Таким образом различные методы применяются совместно – группируются, в зависимости от системы.

Пример: анализаторы сетевого трафика Positive Techologies представляют целые цепочки анализаторов, построенные для максимально эффективного использования вычислительных ресурсов – не забываем, что задача поиска угроз требует затраты, сопоставимые с содержанием такой же системы.

Прогнозирование временных рядов при решении задачи поиска аномалий.

Рекомендательная система как система поиска аномалий.

SNORT

Своя IDS система

Сравнение

Перехват и обработка сетевого трафика.

Типы сетевых атак. Подходы к их выявлению.

Построение алгоритма выявление небезопасной сетевой активности на основе анализа статистических данных по многим факторам.

Реализация гибкой настройки с добавлением своих данных и выбором параметров для возможности использования в узкоспециализированных системах.

https://selectel.ru/blog/ips-and-ids/

<https://www.icmm.ru/uchebnaya-deyatelnost/lektsii/514-ids>

<https://cyberleninka.ru/article/n/obzor-otkrytyh-naborov-dannyh-dlya-vyyavleniya-atak-na-veb-prilozheniya/viewer>

О том, где искать дата сеты

https://habr.com/ru/companies/vasexperts/articles/790192/

Литература

1. Секреты и ложь. Безопасность данных в цифровом мире для блока литература, [Шнайер Брюс](https://readli.net/avtor/SHnayer-Bryus/), 2003
2. <https://selectel.ru/blog/ips-and-ids/>
3. <https://www.securityvision.ru/blog/obnaruzhenie-i-predotvrashchenie-vtorzheniy/>
4. <https://cisoclub.ru/metody-vyjavlenija-i-reagirovanija-na-anomalii-v-kompjuternyh-setjah/>
5. «Временные ряды в прогнозировании спроса, нагрузки на КЦ, товарных рекомендациях и поиске аномалий», [kirillkosolapov](https://habr.com/ru/users/kirillkosolapov/), https://habr.com/ru/articles/477206/
6. «9 подходов для выявления аномалий», [kirillkosolapov](https://habr.com/ru/users/kirillkosolapov/) https://habr.com/ru/articles/477450/

Идеи для реализации

1.Обязательно учесть систему доверенного третьего лица, которое будет реализовывать соответствующий интерфейс.

2.Файловая система – это модуль системы, которых может быть много, это обеспечить высокую пропускную способность, увеличивая нагрузку на сеть

Например: в Mongo DB существуют коллекции, было бы круто равномерно распределить все файлы на несколько «узлов файловых систем», каждый из которых отвечает за свою коллекцию, чтобы можно было параллельно обрабатывать большие потоки данных.

3.Существуют различные системы для хранения файлов на серверах – рассмотреть возможность создания прослойки для работы с такими системами как с будто бы это файловая система нашей машины – это позволит использовать уже существующие решения для создания высокопроизводительной, адаптивной под задачи, системы.

4.Определиться, как рассматривать ОС – пока я представляю ОС как множество сетевых узлов, на которых может быть любая Физическая ОС, подобранная для конкретной функции – т.е. рассматривать ОС как сборку уже готовых решения, на данный момент трудно будет заставить всех перейти на новую структуру

5.В последствии рассмотреть вариант создания своего микроядра, совместимого со всеми аппаратными решениями – это очень сложно(

6.Что делать с графикой? Видеокарта – тоже модуль системы. Полагаю, нужно посылать вычислительные задачи на разные узлы, чтобы успевать обрабатывать запросы графики.

7.Пользовательский компьютер – терминал или очень слабое, в сравнении с системой, устройство для примитивных задач.

8.В данной топологии ОС, которую хочу строить будет мало уязвимостей современных ПК, но не забывать про сетевые атаки – всё взаимодействие идёт по сети, не факт, что локальной.

9.Обязательно систему хранения логов, например, ElasticSearch и системы выявления аномалий(Алгоритмическую и AI)

10.В системе скорее всего будут важные распределительные пункты, они должны быть просты, их цель распределить задачи по распределённым узлам.

11.Поговорить о безопасности(Литература 4 – интересная статья)

12.Сервер реинкарнации – от него порождаются все процессы, он отлавливает подвисшие-зомби процессы, удаляет их, перезапускает и сообщает об этом другим серверам.

13.Возможность склонировать и запустить параллельно уже работающий модуль системы для того, чтобы распараллелить нагрузку или чтобы удобно протестировать без вреда для всей системы.